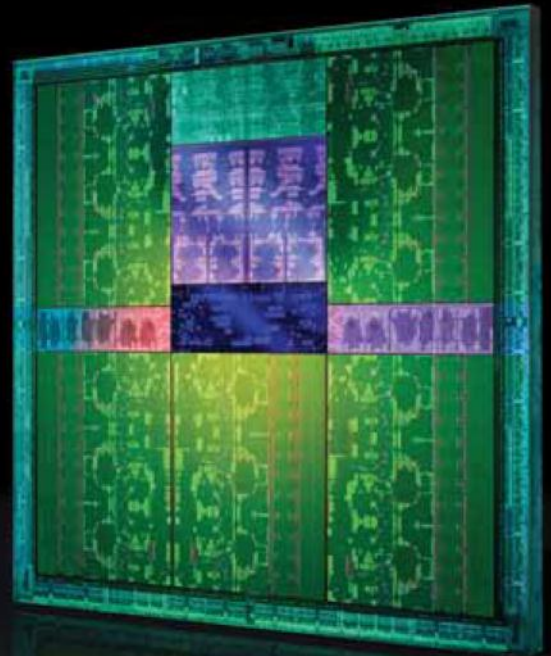




# NVIDIA® KEPLER GK110

## 下個世代的 CUDA® 計算架構



### 最快速、最具效率的 HPC 架構

隨著 2009 年 Fermi GPU 的推出，NVIDIA 迎接高效能計算（HPC）工業的新紀元，其植基於複合式運算模型，其中以 CPU 和 GPU 協同工作解決計算密集的運算工作。

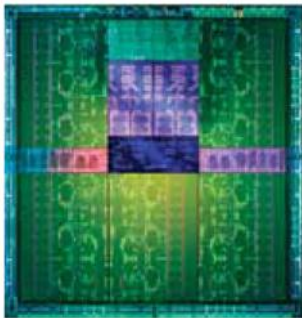


圖 1：Kepler GK110 GPU—世界最快速且最具電力效率的 x86 加速器

在短短幾年間，NVIDIA Fermi GPU 提供計算動力予一些世界上速度最快的超級計算機，以及全球成千上萬的研究團隊。現在，有了新 Kepler GK110 GPU，NVIDIA 又再次地提高了 HPC 工業的門檻。

由 7.1 億個電晶體所組成，Kepler GK110 GPU 是工程奇蹟，係創造用於解決 HPC 中最嚴峻的挑戰。Kepler 係完全從頭開始設計，最大化運算效能而又兼具卓越的電力效率。該架構具有的創新功能可使複合式運算更加容易，適用於更廣泛的應用，且更易於存取。

Kepler GK110 GPU 是一個運算主力，具有每秒兆次的整數、單精度，和雙精度浮點運算效能，以及最高的記憶體頻寬。第一個植基於 GK110 的產品，將是 Tesla K20

GPU 計算加速器。

本技術簡介係設計用於快速地概述，Kepler GK110 GPU 三個最重要的功能：SMX、Dynamic Parallelism，以及 Hyper-Q。額外架構功能的更詳細資料，請參閱 Kepler GK110 白皮書。

#### SMX—次世代的串流多處理器

Kepler GK110 GPU 的核心是新的 SMX 單元，其中包括幾個架構上的創新，其不僅是我們有史以來建構的最強大串流多處理器（SM），也是最具可程式性和電力效率者。

#### DYNAMIC PARALLELISM—創造工作只在眨眼間

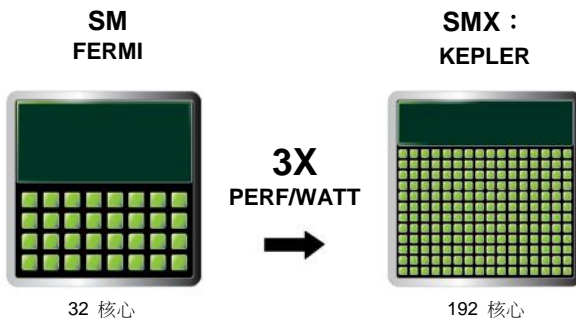


圖 2：SMX：192 個 CUDA 核心、32 個特殊功能單元（SFU），以及 32 存取單元（LD/ST）

## DYNAMIC PARALLELISM

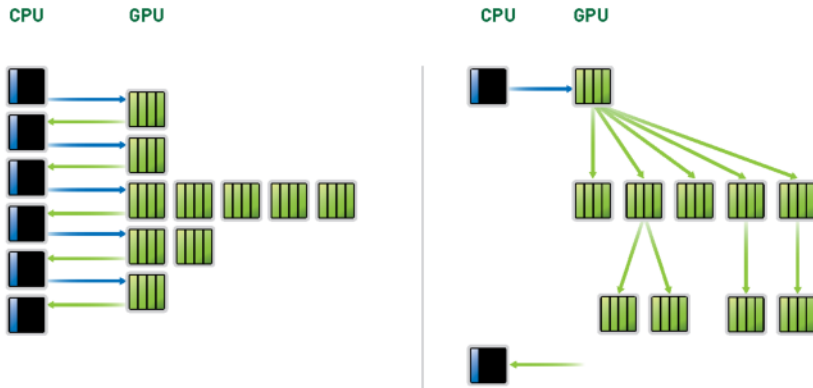


圖 3：在以往，GPU 必須透過 CPU 啟動各個程式核心。利用 Dynamic Parallelism 這項新功能，Kepler GK110 GPU 現在可以獨立啟動嵌套的核心，而無須事先與 CPU 溝通。

設計 Kepler GK110 架構的整體目標之一，是使開發人員能更輕鬆地獲得 GPU 強大平行處理能力的好處。

為此，新的 Dynamic Parallelism 功能使 Kepler GK110 GPU 動態地產生新的執行緒，使之順應於數據而無須返回 CPU。這有效地使更多部分的程式可直接在 GPU 上執行，因為核心現有獨立地依其所需啟動額外的運算工作。

任何一個程式核心可啟動彼此，且可建立必要的串流、事件以及處理額外工作所需要的依存關係，而不需要持續與 CPU 溝通。這簡化的程式設計模型，更易于建立、最佳化與維護。其還藉由保留相同於傳統 CPU 核心啟動的語法，供 GPU 啟動運算工作之用，進而創造了更為友善的程式開發環境。

Dynamic Parallelism 拓寬了應用程式，透過 GPU 於各種領域現在所能完成的事項。應用程式可動態地啟動中小型的平行運算工作，在以前這樣的做法是很昂貴。

### HYPER-Q - 最大化 GPU 的資源

Hyper-Q 讓使用者可透過多個 CPU 在單一 GPU 上同時啟動工作，換句話說就是大幅提高了 GPU 的利用率並有效縮減 CPU 的閒置時間。相較於 Fermi 一對一的情況，此項功能可大幅提升 Kepler GK110 GPU 與 CPU 間的連繫，最多可同時接受 32 個硬體管理的指令。Hyper-Q 是深具彈性的解決方案，該方案允許連線用於 CUDA 串流和訊息傳遞介面 (MPI) 行程，甚至來自一個行程內的工作緒。先前受限於假性依存關係 (false dependencies) 的現有應用程

式，在不改變任何程式碼的情況下，最高可達 32 倍的效能提升。

Hyper-Q 為 MPI 型平行運算系統的使用提供了非常顯著的益處。傳統 MPI 型演算法往往被設計用於多 CPU 系統。由於可經 CPU 系統有效處理的運算工作量，一般小於 GPU 所能提供者，因此送入每個 MPI 行程的運算工作量，不足以完全地佔用 GPU 處理器。

雖然送出多個 MPI 行程，並於 GPU 上同步地執行並非不可能，但是這些行程有可能成為假性依存關係的瓶頸，迫使 GPU 以低於其尖峰效率的狀況下運作。Hyper-Q 移除了假性依存關係瓶頸，並大幅增加從系統 CPU 移往 GPU 供處理之 MPI 行程的速度。

Hyper-Q 自許成為 MPI 應用程式效能的

推手。

### 結論

Kepler GK110 GPU 係設計用於提供突破性的效能，使得 GPU 較以往任何時候都更易於使用，也兼具卓越的電力效率。SMX、Dynamic Parallelism、Hyper-Q 是 Kepler GK110 GPU 三個重要的創新，其將這些好處付諸實現予我們的使用者。有關額外架構功能的更詳細資訊，請造訪 <http://www.nvidia.com/object/nvidia-kepler.html>。參閱 Kepler GK110 白皮書

## NVIDIA HYPER-Q

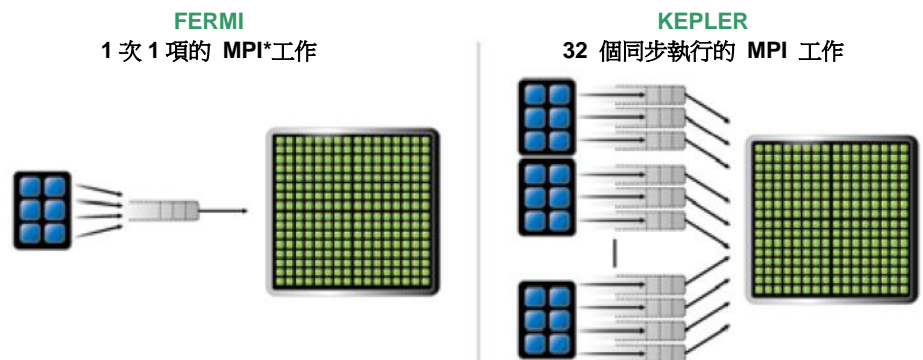


圖 4：Hyper-Q 允許所有的串流，使用不同的工作佇列同步執行。在 Fermi 模型，單一硬體工作佇列所引起的流內 (intra-stream) 依存關係，而使得同步性受到限制。

想要進一步瞭解 NVIDIA Tesla，請造訪 [www.nvidia.com/tesla](http://www.nvidia.com/tesla)